DOCUMENT RESUME

ED 219 411 TM 820 422

AUTHOR Arndt, Stephan
TITLE New Developments in the Analysis and Measurement of Change.
INSTITUTION California Univ., Los Angeles. Mental Retardation Research Center.
SPONS AGENCY National Inst. of Child Health and Human Development (NIH), Bethesda, Md.
PUB DATE [81]
GRANT HD-04612; HD-05540
NOTE 12p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; Analysis of Covariance; Change; Correlation; Longitudinal Studies; *Reliability; *Research Design; *Research Problems; Test Construction
IDENTIFIERS Change Scores; *Growth Curve Analysis; Measurement of Change; *Time Series Analysis

ABSTRACT

The problem of change scores' correlation with initial status and the problem of low reliability in the measurement of change are addressed. By treating the correlation between initial status and change as a design problem rather than a statistical issue, research questions can be formulated in terms of changes in the shapes of growth curves rather than in terms of inter-group differences in change rates. With the recent developments in Time Service Analysis and in the analysis of cross-longitudinal growth curves, the dependent variable remains simple, thus making the interpretation of any shifts in pattern straightforward. The more stable the test measure is over time, the less chance there is of finding a significant effect. From this perspective, only tests which are unstable yet internally consistant will yield sensitive measures of change. If the low reliability is defined as a measurement problem, then it is not something inherent in the change score itself. By re-defining this and other problems associated with measuring change, progress can be made in solving them. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219411

New Developments in the Analysis and Measurement of Change

Stephan Arndt
Psychology Department and
Social Science Training and
Research Laboratory
Memorial Library
University of Notre Dame
Notre Dame, Indiana 46556

Formerly with:
School of Medicine
University of California, Los Angeles
Mental Retardation Research Group
Los Angeles, California 90024

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Arndt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 820 122

This paper was supported in part by Grants No. HD-04612 and HD-05540 from the National Institute of Child Health and Human Development.

New Developments in the Analysis and Measurement of Change

Stephan Arndt

The problems inherent in measuring change have interested researchers for many years now. Difficulties such as low reliability, the high correlation between change scores and initial status, as well as regression to the mean are often encountered and well formalized. These issues are of particular importance to research in mental retardation since almost by definition the developmentally disabled tend to change very slowly. Unfortunately, there are no clear cut and agreed upon solutions to be found in the literature. This lack, I believe, is due to the general perspective that the problems associated with change can be solved using complex statistical analyses. However, more recent work has taken a different perspective and has already yielded promising results. I would like to go through the previous work to illustrate its general thrust and then try to summarize the philosophy of the more recent efforts. First, the problem of change scores' correlation with initial status is discussed since this has received the most attention in the literature. Second, the problem of low reliability in change is addressed. This latter problem is of particular importance to the field of developmental disabilities yet it has received relatively little attention in the methodological literature.

The Correlation Between Initial Status and Change.

As I mentioned before, one problem with change scores is that they are sometimes highly correlated with initial status. This is because the change score is a simple linear combination of the initial and final scores--more precisely, the change score is the post-test minus the

pre-test. Hence the initial score is a part of the change score. As the initial score is subtracted, or given a negative weight, it is usually negatively correlated with the resulting change score. Consequently higher functioning people will tend to change less than lower functioning individuals because of the algebra involved. Since this is frequently thought of as an undesirable characteristic for a measure of growth, Analysis of Covariance is often suggested (e.g., Linn & Slinde, 1977).

With an Analysis of Covariance or ANCOVA, the effects of initial status are statistically removed from the post-test score and groups can be compared for different growth rates. There are several problems with this technique. For one, there have been at least four different suggestions on how to estimate the proper beta weight used to adjust the post-test scores (Bryk & Weisberg, 1977). The normal least squares solution is only applicable for a particular growth pattern. Thus, the researcher must have a sound understanding of the dependent measure's growth curve before choosing the correct beta weight for the covariance analysis. Furthermore, the amount of bias introduced by selecting the wrong weighting factor is sometimes greater than if simple change scores had been used in the first place.

Another problem is that the analysis of covariance has the effect of making all of the groups "hypothetically" equivalent on the pre-test. If the groups were initially different, then pretending that they are equal through statistical adjustments can be extremely misleading (Bock, 1975; Elashoff, 1969; Lord, 1969). What I find to be the most distressing problem with this type of analysis is that the researcher can no longer interpret the results as "one group changed more than another" or that "the changes in the groups were not significantly different". Rather one is

forced to speak in terms of the post-test status once the effects of pre-test status have been removed. Clearly the concept of post-test removing the effects of pre-test is not as simple to interpret or explain as is change (Cronbach & Furby, 1970; Richards, 1975, 1976).

A third problem with the standard Analysis of Covariance is that the dependent measure at pre-test is usually affected by measurement error; thus the pre-test post-test correlation is attenuated due to unreliability (Lord, 1963). A frequently suggested remedy is to correct the correlation matrix using reliability estimates--this is called Analysis of Covariance Corrected for Attenuation. Aside from the difficulties of picking the correct reliability estimate and the number of statistical assumptions now made (Campbell & Boruch, 1975), a more troubling problem exists. We are now a long way away from asking if one group changed more than another. Instead, the dependent measure is the post-test score controlling for the True score component of the initial score (whatever that might be!). In most cases this measure will be at very best extremely difficult to interpret. Even more complex and further removed techniques have been suggested which have been labeled "base-free measure of change" (Cronbach & Furby, 1970; Tucker, Damarin & Messick, 1966). However, I think the pattern is clearly that the statistical analyses get progressively more complex and that the dependent measure quickly becomes modified out of the realm of interpretability.

A New Perspective to the Problem of the Correlation Between Initial Status and Change.

The intercorrelation problem of the change score and initial status is really only a problem when the groups are non-equivalent on the pretest. The other benefits from the Analysis of Covariance, i.e., increased power

and reliability, are usually negligible in the situation where a pre-test is treated as a covariate (Huck & McLean, 1975; Linn & Slinde, 1977). Hence, we are really dealing with a design problem rather than a statistical issue. Much more information on the effects of a treatment over time can be gained by inspecting shifts in the growth curve (Wohlwill, 1973). The more recent developments in Time Series Analysis and in the analysis of cross-longitudinal growth curves (Eyman & Arndt, in press; Nesselroade & Baltes, 1979) should offer many advantages to the ANCOVA techniques. With these more modern strategies the dependent variable remains simple, thus making the interpretation of any shifts in pattern straightforward. One problem in the past has been the accessibility of computer programs that will perform these types of analyses; however, several excellent programs have recently been made available in the BMDP and SAS packages.

Briefly, both of these newer strategies require measurements to be taken at several points in time or development. For Time Series Analysis, longitudinal information is required for all individuals under study and for the duration of the ages of interest. With the cross-longitudinal strategy, people of varying ages are followed for just a few measurement periods and the mini-growth curves are integrated to form a full growth curve. Of importance to both techniques is the shape of the growth curves. Here many of the problems with change scores and ANCOVA are avoided since neither are change scores computed nor is the data adjusted using beta weights. I should add, however, that these strategies do not make any pretense of making non-equivalent groups equivalent.

The Problem of Unreliability of Change Scores.

Sometimes analysis of growth curves is impractical and the researcher would rather phrase his or her questions in terms of simple change. Perhaps the most severe problem here is that change scores for the variables we tend to use are highly unreliable, even when the test at any one time shows good reliability (Linn & Slinde, 1977). For instance, measures of IQ or Adaptive Behavior generally have high reliabilities in the .8 or .9 range, yet change scores based on these measures frequently have reliabilities near zero. In general, the higher the correlation between pre- and post- tests, the lower the reliability of the change score. At the extreme, if the correlation over time is as high as the average reliability of the pre- and post- tests, then the reliability of the change score will be exactly zero no matter what the pre-or post-test's reliability was to begin with. The result of this problem is a drastic reduction in the statistical power of any tests for group differences or correlates of change--the more stable the measure is over time, the less chance there is of finding a significant effect. Using a residualized gain score, one of the "Base-free" measures of growth, or an Analysis of Covariance, approach has very nominal effects at improving this situation--I find that usually the difference is only of one or two points in the second decimal in my work on adaptive behavior. Other researchers have had similar results (St. Pierre, 1978).

As with the previous problem of the correlation between initial status and change, the most fruitful solution is <u>not</u> to treat this as a statistical issue. Quite simply, if the dependent measure, i.e., change, is unreliable there is nothing statistical that can be done. Rather, this is a measurement problem.

Viewed from this perspective, instruments should be constructed so as to have a high change score reliability. Only tests which are unstable yet internally consistant will yield sensitive measures of change. However, I have found that the best approach for test construction is to totally ignore the item statistics for the pre- and post-tests taken individually. Instead, item changes over time should be computed and then these item change scores could be entered into a principal components analysis without rotation. The resulting factor score matrix gives the solution for the linear combination of items having the maximum change score reliability. Frequently, the best items determined by this analysis are not the ones that would be selected using traditional item selection critieria. For instance, of the two general factors in the Behavior Development Survey, Adaptive and Maladaptive Behavior, Adaptive Behavior is far more reliable at any one, time but it is also highly stable (Arndt, 1981). As a result, the change score reliability for Adaptive Behavior is extremely low, generally less than .5. On the other hand, while Maladaptive Behavior tends to have only moderate reliability, it is also unstable; consequently changes in Maladaptive Behavior are far more reliable than changes in Adaptive Behavior.

Selecting items or tests on the basis of their instability while ignoring the information at any one time is not a totally novel procedure (e.g. Bereiter, 1963; Nesselroade, 1977). For instance, Alfred Binet originally selected items which young children tended to fail but as they grew older would tend to pass. Here the item variance at time one or time two would be low and thus, by standard item selection criteria should be excluded, even though the item was a good developmental measure of growth. Clearly, research which is interested in change should use developmentally

sensitive measures and not instruments which were developed with stability of individual differences over time as criteria for item selection. I would like to propose that the low reliability of change scores is not inherent in the use of change scores--rather, the problem lies in our choice of testing instruments. Several lines of research questions become important here, for instance, "Can meausres of adaptive behavior be developed which are sensitive to change?" or "What are the attributes that do change in retarded individuals?"

When selecting already established tests, care should be taken to evaluate the instrument's change score reliability prior to data collection. This can be done so long as internal consistency or parallel forms reliability as well as the test's stability are known. Formulas for calculating the change score reliabilty, given the information on the other reliabilities, can be found in many measurement texts (e.g., Nunnally, 1967). Unfortunately, the number of change-sensitive tests is probably low at present. This problem will continue so long as test makers think of instability as a negative aspect of a test.

To conclude, I think that we are finally making headway in solving the problems associated with change. Basically, this is because we have begun to re-define the problems. By treating the correlation between initial status and change as a design issue we need only formulate our research questions in terms of changes in the shapes of growth curves rather than if one group has changed more than another group. Another issue that has been associated with change scores is unreliability. As I pointed out, if we define the low reliability as a measurement problem, then it is not something inherent in the change score itself but what it is that we actually measure or how the test is constructed that is the problem. While I have

only dealt with work on two of the problems of measuring change, several other issues are showing the same amount of progress. Of course resolutions to the problems are still a ways off, but with further advances solutions may occur in the very near future.

References

Arndt, S. A general measure of adaptive behavior. American Journal of Mental Deficiency, 1981, 85, 554-556.

Bereiter, C. Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Bryk, A. S., & Weisberg, H. I. Use of the nonequivalent control group design when subjects are growing. Psychological Bulletin, 1977, 84, 950-962.

Campbell, D. T., Boruch, R. F. Making the case for randomized assignment to treatments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. A. Bennett & A. A. Lumsdaine (Eds.) Evaluation and experiment. New York: Academic Press, 1975.

Elashoff, J. D. Analysis of covariance: A delicate instrument. American Educational Research Journal, 1969, 3, 383-401.

Eyman, R. K., & Arndt, S. Life-span developement of institutionalized and community based mentally retarded residents. American Journal of Mental Deficiency, in press.

Huck, S. W. & McLean, R. A. Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. Psychological Bulletin, 1975, 82, 511-518.

Linn, R. L. & Slinde, J. A. The determination of the significance of change between pre- and posttesting periods. Review of Educational Research, 1977, 47 (1), 121-150.

Lord, F. M. Elementary models for measuring change. In C. W. Harris (Ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Lord, F. M. Statistical adjustments when comparing pre-existing groups. Psychological Bulletin, 1969, 72, 336-337.

McCleary, R., & Hay, R. A. Applied time series analysis for the social sciences. Beverly Hills: Sage, 1980.

Nesselroade, J. R. Issues in studying developmental change in adults from a multivariate perspective. In J. E. Birren & K. W. Schaie (eds.), Handbook of the psychology of aging, New York: Van Nostrand, 1977.

Nesselroade, J. R. & Baltes, P. B. Longitudinal research in the study of behavior and development. New York: Academic Press, 1979.

Nunnally, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

Richards, J. M. A simulation study of the use of change measures to compare educational programs. American Educational Research Journal, 1975, 12, 199-311.

Richards, J. M. A simulation study comparing procedures for assessing individual educational growth. Journal of Educational Psychology, 1976, 68, 603-612.

St. Pierre, R. G. Correcting covariates for unreliability: Does it lead to differences in an evaluator's conclusions? Evaluational Quarterly, 1978, 2, 401-420.

Tucker, L. R., Damarin, F., & Messick, S. A base-free measure of change. Psychometrika, 1966, 31, 457-473.

Wohlwill, J. F. The study of behavioral development. New York: Academic Press, 1973.